

RESEARCH ARTICLE

Exchange rate predictability and dynamic Bayesian learning

Joscha Beckmann^{1,2} | Gary Koop³ | Dimitris Korobilis⁴ | Rainer Alexander Schüssler^{5,6}

¹Department of Economics, University of Greifswald, Greifswald, Germany

²Kiel Institute for the World Economy, Kiel, Germany

³Department of Economics, University of Strathclyde, Glasgow, UK

⁴Adam Smith Business School, University of Glasgow, Glasgow, UK

⁵Department of Econometrics and Statistics, Technical University of Dortmund, Dortmund, Germany

⁶Department of Economics, University of Rostock, Rostock, Germany

Correspondence

Gary Koop, Department of Economics, University of Strathclyde, 199 Cathedral Street, Glasgow G4 0QU, UK.
Email: gary.koop@strath.ac.uk

Summary

We consider how an investor in the foreign exchange market can exploit predictive information by means of flexible Bayesian inference. Using a variety of vector autoregressive models, the investor is able, each period, to learn about important data features. The developed methodology synthesizes a wide array of established approaches for modeling exchange rate dynamics. In a thorough investigation of monthly exchange rate predictability for 10 countries, we find that using the proposed methodology for dynamic asset allocation achieves substantial economic gains out of sample. In particular, we find evidence for sparsity, fast model switching, and exploitation of the exchange rate cross-section.

1 | INTRODUCTION

Understanding and predicting the evolution of exchange rates has long been a key component of the research agenda in international economics and finance. Yet, the early finding by Meese and Rogoff (1983) that structural models cannot offer predictability superior to that of a random walk has not been convincingly overturned. The voluminous existing literature on exchange rate forecasting, surveyed in Rossi (2013), adopts many different econometric methods. Broadly speaking, these differences fall into the following categories. First, they differ in whether they are multivariate (e.g., building a vector autoregressive (VAR) model involving a cross-section of exchange rates for many countries) or univariate. Second, they differ in which predictors they use. Third, they differ in how they treat the fact that there may be many potential predictors, most of which are unimportant. Fourth, they differ in whether they allow for dynamic model change (i.e., whether the best forecasting model can involve different predictors at different points in time) or not. Fifth, they differ in whether they allow for parameter change (both in VAR or regression coefficients and in volatilities) or not.

We develop an econometric approach that allows for a general treatment of each of these five categories. That is, its most flexible specification is a high-dimensional multivariate time series model involving the full cross-section of exchange rates, several exogenous predictors, and time variation in coefficients and volatilities. Our algorithm, however, allows for decisions relating to these categories to be made in a data-based fashion using dynamic model selection methods. That is, the estimation procedure automatically decides whether to set a coefficient on a predictor or a VAR lag to be zero (or not). Most importantly, it does so in a dynamic manner, allowing for different forecasting models to be used at different

points in time. Thus decisions about specification choices (i.e., different predictors, different VARs, different degrees of model switching) are all made automatically in a time-varying fashion.

Our econometric approach is related to papers such as Koop and Korobilis (2013) and Giannone, Lenza, and Primiceri (2015), which provide strategies for handling prior elicitation and dynamic uncertainty in VAR models. We improve on and extend them in important directions of relevance for our empirical application. These include, in particular, rich shrinkage patterns provided by a Minnesota-type prior and flexible treatment of exogenous regressors.

Our framework enables us to assess the relative contributions of different modeling aspects in an exchange rate forecasting exercise involving 10 countries and exogenous predictors. We take the view of a Bayesian investor with a broad perspective, accommodating many features inspired by the exchange rate literature. Hence our focus is on the economic evaluation of the density forecasts generated by our approach. That said, given the wealth of empirical results provided as a byproduct, we explore them to relate our findings to previously documented characteristics and phenomena of exchange rate behavior.

To preview our empirical results, we do find that model switching has a big role to play. At most points in time only one or a few predictors are relevant for forecasting. In general, the best economic results are achieved when both VAR lags and fundamentals are considered in the candidate models, though we find that VAR lags and fundamentals act as substitutes to a certain extent in this regard. There are also several periods, however, where a simple multivariate random walk with stochastic volatility is the best forecasting model. We find that an investor using our algorithm would experience substantial economic gains out of sample relative to the random walk model with time-varying volatility. A risk-averse mean-variance investor is willing to pay an annualized fee of several hundred basis points (after transaction costs) for switching from the dynamic portfolio strategy implied by the random walk with constant volatility model to the dynamic asset allocation implied by our VAR-based approach. Similarly, we find that the annualized Sharpe ratio after transaction costs increases substantially from adopting our approach.

The remainder of the paper is organized as follows. Section 2 relates our modeling strategy to the literature. Section 3 discusses the data, and Section 4 lays out our econometric methods. Section 5 presents and discusses our empirical results and Section 6 concludes. We present technical details of our econometric methods along with many empirical results and further details regarding the underlying data in a Supporting Information Appendix.

2 | RELATION TO THE LITERATURE

We provide a selective review of the literature, with a focus on the five econometric modeling issues described in the Introduction.¹ A large part of the existing literature relies on macroeconomic fundamentals to forecast exchange rates with little success, which is commonly referred to as the exchange rate disconnect puzzle.

The scapegoat approach of Bacchetta and van Wincoop (2004) attributes this failure to the fact that market participants attach excessive weight to observable fundamentals that deviate from their long-run trend. As a result, agents quickly switch between models over time and different fundamentals may be relevant only for short periods. This explanation translates into an econometric model that should allow for the optimal forecasting model to change over time.²

Next is the issue of whether parameter change and other nonlinearities are beneficial for forecasting. Overall, the evidence is not strong (Rossi, 2013), although several studies find some benefits from allowing for time variation in parameters; see, for example, Rossi (2006) or Byrne, Korobilis, and Ribeiro (2016). Our approach accommodates both constant and time-varying parameters.

The question of whether there are benefits in working with a multivariate time series model such as a VAR involving a cross-section of exchange rates is also debated. Such an approach has the advantage that it exploits information in the comovements and common dynamics in exchange rates. There is some evidence that doing so can improve exchange rate forecasts. Carriero, Kapetanios, and Marcellino (2009) work with a large Bayesian VAR involving a cross-section of exchange rates and find forecast improvements from considering dynamic comovements of exchange rates. Abbate and Marcellino (2018) extend Carriero et al. (2009) by allowing for, among other things, time-varying coefficients and volatilities, and find the latter to be particularly useful in improving forecast performance. These considerations suggest that working with VARs with time-varying volatilities is potentially important and our modeling approach does so.

¹A thorough review of the voluminous literature on exchange rate predictability can be found in Rossi (2013).

²This *dynamic model switching* aspect has also been found to be of crucial importance in the empirical exchange rate literature. Sarno and Valente (2009) discuss how the fact that there is evidence of a weak link between in-sample fit and out-of-sample predictability complicates the choice of selecting an appropriate model even if fundamentals contain valuable information about the path of the exchange rate.

Another issue that arises when we have many potential predictors is the need for some method for ensuring parsimony so as to avoid overfitting and poor out-of-sample results. Indeed, even in univariate models, papers such as Ackermann, Pohl, and Schmedders (2016) find parameter estimation error to be substantial and hence they use no predictors when building a diversified FX portfolio. Instead they focus solely on exploiting volatility timing. However, several recent papers have used data reduction methods, priors, or model averaging methods to minimize overfitting concerns.

Other techniques have been successfully used, including elastic net shrinkage (Li, Tsiakas, & Wang, 2015), gradient boosting (Berge, 2014), and model averaging/selection (Della Corte, Sarno, & Tsiakas, 2008; Della Corte & Tsiakas, 2012; Kouwenberg, Markiewicz, Verhoeks, & Zwinkels, 2017). All these approaches find sparsity to be an important modeling feature and, in particular, Kouwenberg et al. (2017) illustrate also the time-varying relevance of regressors in a univariate framework. Our work corroborates these findings in a multivariate approach that allows us to assess the incremental value of fundamentals in addition to VAR lags and vice versa. From an investor's point of view, our multivariate approach allows for directly mapping the (density) forecasts into portfolio weights without having to rely on additional procedures as is the case for univariate approaches.

The literature has also explored the implications of exchange rate predictability (or a lack thereof) for an investor wishing to build an investment portfolio involving various exchange rates; see, for instance, Abhyankar, Sarno, and Valente (2005), Della Corte et al. (2008), Kouwenberg et al. (2017), and Abbate and Marcellino (2018).

Motivated by these considerations, our econometric approach takes the perspective of an investor who learns from past mistakes. We formalize this setting econometrically using the notion of dynamic Bayesian learning. In it, the investor can adapt to a new forecasting environment each time period by switching to a new model. The decision to switch is based on past forecast errors. The result is an extremely flexible framework that learns quickly from recent forecast performance. Our empirical framework has several desirable features. First, due to the specification of time-varying parameters and dynamic model switching, the VAR forecasting model can adapt to abrupt structural changes or sudden shifts in the investor's information set. Our estimation methods are Bayesian so that the investor's decisions account for parameter uncertainty. At the same time, Bayesian methods offer a natural setting for imposing statistical shrinkage which, as discussed above, has been shown to be important for exchange rate predictability when working with large numbers of predictors and a large cross-section of exchange rates. Finally, it is worth mentioning that we allow for model incompleteness; see, for instance, Billio, Casarin, Ravazzolo, and van Dijk (2013). That is, we do not assume that one of our entertained VARs reflects the correct data-generating process. The Supporting Information Appendix includes a small simulation experiment that outlines how model incompleteness is accommodated.

3 | DATA

All of our individual model configurations are VARs (or extensions thereof) that involve a cross-section of exchange rates as dependent variables. Some models also include additional exogenous predictors. We use the common set of G10 currencies: the Australian dollar (AUD), the Canadian dollar (CAD), the euro (EUR),³ the Japanese yen (JPY), the New Zealand dollar (NZD), the Norwegian krone (NOK), the Swedish krona (SWK), the Swiss franc (SWF), the Great Britain pound sterling (GBP) and the US dollar (USD). All currencies are expressed in terms of the US dollar and are end-of-month exchange rates that enter the model as discrete returns. Thus we have nine exchange rates, each relative to the US dollar, entering our VAR. The sample runs from 1986:01 until 2016:12. As additional predictors, we also include the uncovered interest parity (UIP), the percentage change in stock prices over the past 12 months (STOCK GROWTH), the difference between long- and short-term interest rates (INT DIFF), and the percentage change in the nominal oil price (OIL). UIP, STOCK GROWTH, and INT DIFF have been widely used in studies such as Wright (2008), and previous research shows that US dollar exchange rates are affected by the price of oil (Lizardo & Mollick, 2010). With regard to the interest rates, we use 1-month LIBOR and Eurodeposit interest rates as well as 10-year government bonds.

In the Supporting Information Appendix, we present empirical results with a longer sample of data going back to 1973 and also provide results for additional established predictors, such as purchasing power parity, the monetary model and the Taylor rule approach, which are not available in real time. These results mainly reinforce the findings presented below. We focus on the shorter sample since it covers a period where all exchange rates are largely freely floating and availability of predictors is not a concern. The 1970s included several periods of economic turbulence, such as the oil price shock and changes in exchange rate arrangements for some currencies such as Sweden or Norway. In addition, perfectly comparable interest rates are not available.

³We use German instead of euro data prior to 1999.

The forecast evaluation period runs from 1996:01 to 2016:12 for a total of 252 observations. The Supporting Information Appendix provides sources, descriptions of the fundamental exchange rate models, other details about the data, and results for the long sample.

4 | ECONOMETRIC METHODS

4.1 | The VAR

Our starting point is a time-varying parameter VAR with exogenous variables:

$$\mathbf{y}_t = \mathbf{x}_t \boldsymbol{\beta}_t + \varepsilon_t, \quad \varepsilon_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_t), \quad (1)$$

$$\boldsymbol{\beta}_{t+1} = \boldsymbol{\beta}_t + \mathbf{u}_t, \quad \mathbf{u}_t \sim N(\mathbf{0}, \boldsymbol{\Omega}_t), \quad (2)$$

where \mathbf{y}_t is an $M \times 1$ vector containing observations on M time series variables (in our case, discrete exchange-rate returns for nine countries). \mathbf{x}_t is a matrix where each row contains predetermined variables in each VAR equation, namely an intercept, (lagged) exogenous variables, and p lags of each of the M variables. We divide the set of exogenous variables into two groups: N_x denotes the number of variables which are asset specific and considered as relevant only for a specific exchange rate. For instance, in the equation for the UK currency the UIP for the UK belongs in this class. N_{xx} denotes the number of non-asset-specific variables which are supposed to be potentially relevant for all currencies in the setting (e.g., oil price changes). Thus, we have $k = M(1 + p \cdot M + N_x + N_{xx})$ elements in $\boldsymbol{\beta}_t$. Following a large literature in economics and finance,⁴ we assume that $\boldsymbol{\beta}_t$ evolves as a multivariate random walk without drift, with covariance matrix $\boldsymbol{\Omega}_t$ of dimension $k \times k$.

Here we outline our methods for estimating and forecasting with a single VAR. Additional details are given in the Supporting Information Appendix. We require a prior for the initial condition for the time-varying VAR coefficients. In the case of the constant-coefficient VAR, this is the prior for the VAR coefficients. We use a variant of the Minnesota prior:⁵

$$\boldsymbol{\beta}_0 \sim N(\mathbf{0}, \boldsymbol{\Omega}_0).$$

Hence model coefficients are initialized with an expected value of 0 and covariance matrix $\boldsymbol{\Omega}_0$. If the diagonal elements of $\boldsymbol{\Omega}_0$ are chosen to be small, the respective coefficients are shrunk to 0. We employ this mechanism to effectively exclude certain exogenous variables in some model configurations. The Minnesota prior assumes the prior covariance matrix $\boldsymbol{\Omega}_0$ to be diagonal. Let $\boldsymbol{\Omega}_{0,i}$ denote the block of $\boldsymbol{\Omega}_0$ associated with the coefficients in equation i and $\boldsymbol{\Omega}_{0,i,jj}$ its diagonal elements. The shrinkage intensity towards 0 is determined by the hyperparameters γ . We assume a prior covariance matrix of the form

$$\boldsymbol{\Omega}_{0,i,jj} = \begin{cases} s_i^2 \gamma_1 & \text{for intercepts} \\ \frac{\gamma_2}{r^2} & \text{for coefficients on own lag } r = 1, \dots, p \\ \frac{\gamma_3 s_i^2}{r^2 s_j^2} & \text{for coefficients on lag } r \text{ of variable } i \neq j \text{ for } r = 1, \dots, p \\ \gamma_4 s_i^2 & \text{for coefficients on the first asset-specific exogenous variable} \\ & \text{associated with exchange rate } i \\ \dots & \dots \\ \gamma_{N_x+3} s_i^2 & \text{for coefficients on the last asset-specific exogenous variable} \\ & \text{associated with exchange rate } i \\ \gamma_{N_x+4} s_i^2 & \text{for coefficients on the first non-asset-specific exogenous variable} \\ \dots & \dots \\ \gamma_{N_x+N_{xx}+3} s_i^2 & \text{for coefficients on the last non-asset-specific exogenous variable,} \end{cases} \quad (3)$$

where s_i^2 denotes the residual variance of the respective variable i . We set lag length $p = 6$.

⁴See, for example, Byrne et al. (2016) or Dangl and Halling (2012).

⁵This is the most popular prior for Bayesian VARs, with Banbura, Giannone, and Reichlin (2010) being an early example of its use with large Bayesian VARs, and Koop and Korobilis (2013) using it with large TVP-VARs.

The Minnesota prior is typically controlled by a single shrinkage parameter; see Bańbura, Giannone, and Reichlin (2010) and citations therein. In order to deal with prior sensitivity associated with selecting a particular value for this shrinkage parameter, Giannone et al. (2015) and Koop and Korobilis (2013) use information in the data to learn about its value. We adopt a similar approach and allow the degree of shrinkage in the Minnesota prior to adaptively change over time. Furthermore, by using the prior covariance matrix specified in Equation 3 we allow for richer shrinkage patterns. Instead of having one shrinkage parameter for all VAR coefficients, we allow for multiple shrinkage parameters. For γ_2 and γ_3 we consider values taken from the grid $\{0; 0.1; 0.5; 0.9\}$. We have an intercept, $N_x = 3$ asset-specific exogenous variables (UIP, STOCK GROWTH, and INT DIFF) and $N_{xx} = 1$ non-asset-specific exogenous variable (OIL). We standardize the exogenous variables in real time to have means 0 and variances 1. For all of these, we use a grid of $\{0; 1\}$. All of these choices for grids for γ_i reflect a desire to allow for the algorithm to select either a 0 (which means that the i th variable or block of variables is omitted from the model) or less informative prior choices. Hence our method allows for the exclusion of model elements such as VAR lags or predictors if this is empirically warranted.

Assuming Σ_t and Ω_t are known and the prior for β_0 is as above, standard Bayesian methods for state-space models involving the Kalman filter can be used to estimate β_t and obtain the predictive distribution of the returns.

In practice, the econometrician/investor does not observe Σ_t and Ω_t . In small models, these parameters can be estimated with Markov chain Monte Carlo (MCMC) methods using approaches such as Chib, Nardari, and Shephard (2006). However, when working with larger models MCMC methods become too computationally demanding. Accordingly, we rely on exponential discounting methods. These are filtering methods in which Σ_t and Ω_t are updated by looking at recent data and discounting more distant observations at a higher rate. Thus, if an abrupt change occurs, parameter estimates can adapt at a faster rate compared to an investor who tracks parameters based on the whole, equally weighted, sample of data.

The mechanics behind the discounting approach are described in the Supporting Information Appendix.⁶ The key point to note here is that they involve the use of discount factors δ and λ to control the dynamics of Σ_t and Ω_t , respectively. These two discount factors control how quickly/slowly investors learn from past forecasting performance. When $\delta = 1$ (similarly for λ), then the investor uses all available historical observations, equally weighted, to update volatilities and parameters. For values less than one, older observations are exponentially penalized, giving more weight to recent observations. As we work with monthly data, we set $\delta = 0.97$, following JP Morgan/Reuters (1996) and, hence, allowing for time-varying observational volatilities and covariances. For our main results we select constant slope coefficients, setting $\lambda = 1$, and investigate the effects of time-varying slope parameters ($\lambda = 0.99$) as a robustness check in our Supporting Information Appendix. The choice of constant slope parameters for our main results is motivated by the empirical finding that in medium-sized VARs such as ours it is common to find strong evidence for time-varying error variances, but little evidence in favor of time-varying VAR coefficients (Chan & Eisenstat, 2018; Koop & Korobilis, 2013). Time-varying VAR coefficients may be even detrimental for portfolio performance in the case of FX portfolios (Abbate & Marcellino, 2018). Using Wishart Matrix discounting, we rely on a fully Bayesian approach for modeling the uncertainty surrounded by point forecasts of volatilities and correlations. In our Supporting Information Appendix we provide point forecasts of volatilities and correlations with credibility intervals.

In sum, Bayesian posterior and predictive inference of a single VAR can be done using standard Kalman filtering and discounting methods. A VAR is defined by making particular choices for $\delta, \lambda, \gamma_1, \dots, \gamma_7$ to a particular value. Our proposed structure of the prior can also be motivated as a spike-and-slab prior, a perspective we outline in the Supporting Information Appendix.

4.2 | Dynamic model learning

Our empirical results fix δ and λ and consider a grid of values for each of $\gamma_1, \dots, \gamma_7$ to allow for variable exclusion and different degrees of shrinkage intensity. If we consider every possible combination of values taken from all of these grids we have 512 choices. We interpret a choice as defining a model that the investor has at their disposal at each point in time upon which they could base their portfolio allocation. In order to allow for the investor to make an optimal choice each period t , we use the notion of dynamic model learning (DML). Dynamic model learning involves selecting, at each point in time, the model specification with the highest discounted joint log-predictive likelihood at that time. The predictive likelihood is a measure of out-of-sample forecasting ability that takes into account the entire predictive distribution; see Geweke

⁶Discount factors are well established; see the JP Morgan/Reuters (1996) Riskmetrics model and Dangl and Halling (2012) for an application in stock return predictability. For a general treatment, see West and Harrison (1997).

and Amisano (2012). The individual model configuration with the highest discounted joint log-predictive likelihood is used in order to obtain the predictive mean and covariance matrix. These are a crucial input in portfolio optimization. Our motivation for using learning based on past forecast performance is that it potentially allows for a different model at each point in time. Such a feature is likely particularly useful in times of abrupt change. If we were to use a single VAR, gradual parameter changes would be accommodated if the discount factors δ and λ were below one. But this is not the same as switching between entirely different models as dynamic model learning allows for.

In this dynamic model learning setting, the discounted joint predictive likelihood (*DPL*) can be calculated as

$$DPL_{t|t-1,j} = \prod_{i=1}^{t-1} [p_j(y_{t-i}|y^{t-i-1})]^\alpha,$$

where $p_j(y_{t-i}|y^{t-i-1})$ denotes the predictive likelihood of model j in period i and $t|t-1$ subscripts refer to estimates made of time- t quantities given information available at time $t-1$. Hence model j will receive a higher value at a given point in time if it has forecast well in the recent past, using the predictive likelihood (i.e., the predictive density evaluated at the actual outcome) as the evaluation criterion. The interpretation of “recent past” is controlled by the discount factor α , reflecting exponential decay. For example, if $\alpha = 0.95$, forecast performance 3 years ago receives approximately 15% as much weight as the forecast performance last period. If $\alpha = 0.90$, then forecast performance 3 years ago receives only about 2% as much weight. The case $\alpha = 1$ implies no discounting and the discounted predictive likelihood is then proportional to the marginal likelihood. Lower values of α are associated with more rapid switching between models. We consider a range of values for α and, at each point in time, choose the best value for it. In this way, we can allow for times of fast model switching and times of slow model switching.

At time τ , we choose the best value for α as the one which has produced the model with the highest product of predictive likelihoods⁷ in the past from $t = 1, \dots, \tau$. We consider the following grid of values: $\alpha \in \{0.50; 0.70; 0.80; 0.90; 0.99; 1\}$.

5 | EMPIRICAL RESULTS

5.1 | Evidence on model switching and sparsity

Our most flexible approach allows for dynamic model learning over a set of 512 different VAR models and six different values of α using the methods described in Section 4. We use the term “DML with ALL REGRESSORS” to denote the case where DML is being done over all specification choices including all of the exogenous predictors. “DML without own/cross-lags and NO REGRESSORS” is the (heteroskedastic) random walk. We also consider several restricted versions of DML which involves dynamic model learning over only some of the predictors. “DML with OIL”, for example, means that OIL is the only possible exogenous variable which could be chosen. “DML without cross-lags” means that the coefficients on the cross-lags are set to zero in all VARs. We implement such restrictions by tuning the vector of shrinkage parameters $\gamma_1, \dots, \gamma_7$ introduced in Equation 3. For instance, to delete the effect of cross-lags we set γ_3 to zero. The label DML denotes the VAR which involves only exchange rates (no exogenous predictors). We also consider versions of our approach which set α to a specific value. For instance, DML ($\alpha = 0.99$) means that α is fixed at 0.99 rather than being selected from a grid of values.

The main focus of this paper is on how well these specifications perform in terms of our dynamic asset allocation problem. However, before doing this, we present a few results illustrating how the dynamic model learning strategy is working using the most flexible specification.

Dynamic model learning is to be preferred over static Bayesian model learning only if the optimal forecasting model is changing over time. Figure 1 shows that it does so in our application. The vertical axis plots the model numbers from 1 to 512 against time for two cases. The set of models begins with model number 1, which is the multivariate random walk without drift, and ends with model number 512, which is one of the most flexible models (i.e., the VAR model with an intercept, own lags with shrinkage parameter $\gamma_2 = 0.9$, and cross-lags with shrinkage parameter $\gamma_3 = 0.9$, and with inclusion of all exogenous regressors).

The two lines in Figure 1 are for DML with ALL REGRESSORS (with α selected in a time-varying manner) and DML with ALL REGRESSORS (with α fixed to 1). The latter can be interpreted as allowing for model learning, but using conven-

⁷We stress that we are not using the *DPL* when choosing between different values for α . The *DPL* is only used to select the best model for a given value of α .

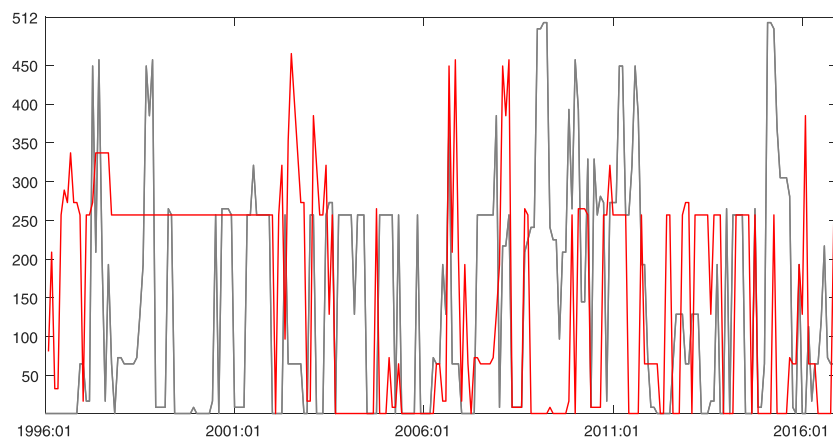


FIGURE 1 Frequency of model change. The vertical axis represents the model configurations 1, ..., 512. The red line depicts the evolution of the selected model configuration for $\alpha = 1$. The gray line shows the evolution of the selected model configuration when α is dynamically chosen from the grid of values $\alpha \in \{0.50; 0.70; 0.8; 0.90; 0.99; 1\}$ [Colour figure can be viewed at wileyonlinelibrary.com]

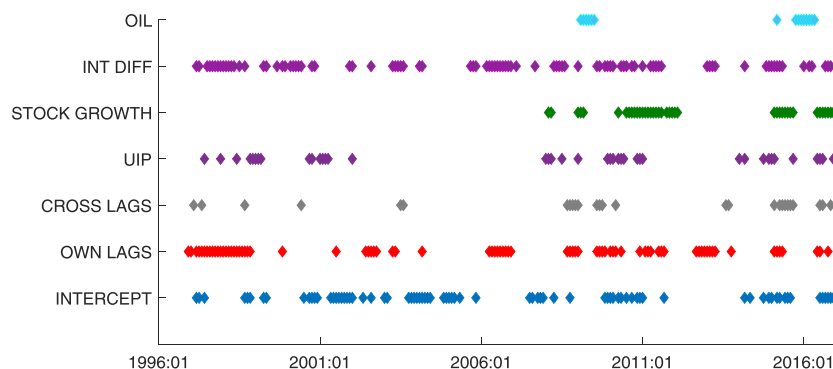


FIGURE 2 Inclusion of blocks of variables. The figure displays which blocks of variables are included at each point in time. “Included” means the respective γ_i is not 0 [Colour figure can be viewed at wileyonlinelibrary.com]

tional Bayesian model averaging methods. Both cases show that different models are selected at different times. However, with our flexible specification where α is chosen in real time, the model change is dramatic, suggesting that a high degree of model switching is a crucial feature. A wide range of different models is selected, with none emerging as dominant. Interestingly, the multivariate random walk is selected 30.16% of the time. It is also evident that in certain episodes in time—for example, during the subprime crisis—flexible models are preferred.

The colored diamonds in Figure 2 show which blocks of variables are included at each point in time. In contrast, blank spaces in the graph depict the time-varying sparsity induced by DML—that is, periods where a block of variables is not selected. There is no single block of variables selected in all periods; however, in most cases, selection of a block persists for several consecutive months before it becomes again irrelevant.

5.2 | Evaluation of economic utility and forecast performance

The previous subsection established that the DML approach is picking up model change, but we have not provided evidence whether this feature is relevant for dynamic portfolio choice. To investigate this further, we design an international asset allocation strategy that involves trading the US dollar and nine other currencies. We consider a US investor who builds a portfolio by allocating their wealth between 10 bonds: one domestic (US), and the nine foreign bonds. In each period, the foreign bonds yield a riskless return in the local currency and at the same time a risky return that is due to currency fluctuations relative to the US dollar. Therefore, the only risk the US investor is exposed to is foreign exchange risk. Every period the investor takes two steps. First, they use the currently selected model (i.e., the model with the highest discounted sum of predictive likelihoods) to forecast the one-period-ahead exchange rate returns and the predictive covariance matrix. Second, using these predictions, they dynamically rebalance their portfolio by calculating the new optimal weights. This setup is designed to assess the economic value of exchange rate predictability and to dissect which sources of information are valuable for asset allocation.

The dynamic asset allocation strategy is described in detail in the Supporting Information Appendix. It involves choosing the investor's degree of relative risk aversion θ . We set $\theta = 2$ and also consider $\theta = 6$ in the Appendix an additional robustness check. It also takes into account transaction costs, τ , ex ante (i.e., at the time of the portfolio construction). Following Della Corte and Tsiakas (2012), we set $\tau = 0.0008$. It also involves choosing a target portfolio volatility, σ_p^* ,

TABLE 1 Evaluation of forecasting results

	Φ^{TC}	SR	SR^{TC}	PLL
DML with UIP	464*	1.12**	0.93**	22.01*
<i>Alternative sets of regressors</i>				
DML with OIL	199	0.89	0.70	22.03*
DML with INT_DIFF	388*	1.06*	0.88*	22.02*
DML with STOCK_GROWTH	368*	1.06*	0.88*	22.06*
DML with ALL REGRESSORS	397*	1.02*	0.87*	22.04*
DML with NO REGRESSORS	327	1.01*	0.82*	22.02*
<i>Type of restrictions: VAR lags</i>				
DML without own lags ($\gamma_2 = 0$) and NO REGRESSORS	98	0.72	0.60	21.97
DML without cross-lags ($\gamma_3 = 0$) and NO REGRESSORS	200	0.86	0.79	21.78**
DML without own/cross-lags ($\gamma_2 = \gamma_3 = 0$) and NO REGRESSORS	5	0.54	0.53	21.72**
DML without own/cross-lags ($\gamma_2 = \gamma_3 = 0$) but with ALL REGRESSORS	255	0.80	0.73	22.00*
<i>Type of restrictions: model selection dynamics</i>				
$\alpha = 1$	-427	0.34	0.11	21.69
$\alpha = 0.99$	-464	0.28	0.08	21.66
$\alpha = 0.90$	98	0.77	0.60	21.96*
$\alpha = 0.80$	266	0.94	0.76	22.02*
$\alpha = 0.70$	327	1.01*	0.82*	22.02*
$\alpha = 0.50$	84	0.82	0.60	21.98

Note. The table summarizes the economic and statistical evaluation of our forecasts from different model configurations for the period from 1996:01 to 2016:12. We measure statistical significance for differences in performance fees (Φ^{TC}) and joint predictive log-likelihoods (PLLs) based on the (one-sided) Diebold and Mariano (1995) *t*-test using heteroskedasticity and autocorrelation robust (HAC) standard errors. We evaluate whether the Sharpe ratio before/after transaction costs (SR/SR^{TC}) of a model is different from that of the random walk (with constant volatility) benchmark using the (one-sided version of the) Ledoit and Wolf (2008) bootstrap test. We compute the Ledoit and Wolf 2008 test statistics with a serial correlation-robust variance, using the prewhitened quadratic spectral estimator of Andrews and Monahan (1992). Asterisks indicate significance at the *10%, **5% level, and ***1% level. Restrictions on α correspond to the specification DML with NO REGRESSORS.

which we set to 10%. We assess the economic value of different forecasting approaches by equating the utility generated by a portfolio strategy that is based on our approach and the utility achieved by a portfolio strategy relying on a simple random walk. The annualized performance fee an investor is willing to pay to switch from a homoskedastic multivariate random walk to our approach is labeled Φ^{TC} in Table 1. As an additional measure of economic utility, we report the Sharpe ratio before and after transaction costs, SR and SR^{TC} (benchmarked relative to the random walk).

The statistical criteria we use are the average joint predictive log-likelihood (PLL), coverage statistics of interval forecasts and the mean squared forecasting error (MSFE). We report (PLL)-statistics in Table 1 and statistics of interval forecasts along with the mean squared forecasting error relative to the random walk in our Supporting Information Appendix. Table 1 contains the results using our approach and the various restricted versions of it described above.

Using DML we find that the annualized performance fee after transaction costs is 327 basis points, and the annualized Sharpe ratio is 1.01 before transaction costs and 0.82 after transaction costs. Including exogenous regressors in DML leads to substantially stronger improvements when using the economic evaluation criteria. For instance, the annualized performance fee after transactions costs increases to 397 when all the regressors are considered. Among the exogenous regressors, including UIP leads to the largest improvements in the economic performance measures. However, with the exception of OIL, the other regressors also lead to improvements. Importance of VAR lags is identified at some points in time and neglecting own lags or cross-lags (i.e., setting $\gamma_2/\gamma_3 = 0$) is detrimental for portfolio performance. These patterns are in line with those in Figure 2. Most of these findings are statistically significant relative to the homoskedastic multivariate random walk benchmark. As noted above, we also repeated this exercise using a longer sample going back to 1973. Using the same economic and statistical criteria, results for this longer sample (reported in the Supporting Information Appendix) are qualitatively similar.

In terms of economic utility gains, our DML models compare very well to results reported in the literature. Given our long evaluation period (252 observations in the “short” sample used in the main text and 324 in the “long” sample reported in the Supporting Information Appendix) and robustness to alternative specifications, this is good news for an investor. However, it is also part of the story that the multivariate models we use involve estimating additional parameters relative to univariate approaches. Enforcing stronger sparsity, we achieve lower mean squared errors but also lower PLLs and

economic gains as a result of narrowing the model space to less parameter-rich configurations. This finding aligns with the results of Cenesizoglu and Timmermann (2012), who report broad agreement between density forecast measures and economic performance measures based on the predictive density. At the same time, they note that there is typically a weak link between point forecast evaluation criteria and economic evaluation criteria. Along these lines, this kind of disagreement in forecasting measures is not uncommon and has also been documented for exchange rates (see, e.g., Abbate & Marcellino, 2018; Fratzscher, Rime, Sarno, & Zinna, 2015).⁸ An exception in this respect is Kouwenberg et al. (2017), who report successful results also in terms of point forecasting accuracy, though at a quarterly data frequency over a different evaluation period and exchange rates/fundamentals different from ours. We report results for our sample based on their approach in the Supporting Information Appendix.

From a statistical perspective it is worth noting that PLLs indicate that if one had to select one single best model for the entire evaluation period, this would have been the multivariate random walk with time-varying error covariance. The corresponding value for the PLL is 21.72 and slightly above 22 for the DML configurations. DML, which allows for the selected model to change over time, would choose the multivariate random walk with time-varying error covariance roughly one third of the time, and leads to higher PLLs and substantially higher economic performance. This demonstrates that our flexible learning mechanism is able to efficiently switch between different model configurations in real time. Assessment based on the results of the long sample (see the Supporting Information Appendix) is comparable.

We next delineate the effect of restrictions on α . In practice, we estimate α to change rapidly over time. The results in Table 1 relating to α show the benefits of this for forecasting. Fixing $\alpha = 1$ (or to other high values) rather than choosing the value of α in real time leads to very poor forecasting results. Allowing for lower values of α and thus more model switching leads to higher values of the log scores and, in particular, to higher performance fees and Sharpe ratios. In fact, the highest performance fee and Sharpe ratio is obtained when $\alpha = 0.70$ for all presented model configurations. Note that the restrictions on α in Table 1 are provided for DML with NO REGRESSORS.

Thus large economic and statistical losses occur if the investor does not emphasize the most recent forecast performance when selecting the forecasting model on which to base their asset allocation decision. Altogether, we are finding the choice of the discount factor α to be a very important one.

The Supporting Information Appendix contains more empirical evidence that expands on and reinforces the story told above. That is, as an econometric approach DML is performing well and, if used to construct an investment portfolio, would yield higher levels of utility than a simple benchmark. In particular, it presents results which show that the coverage of our predictive densities is good and carries out a variety of robustness checks, including alternative prior specifications and the use of different sets of predictors. We find the modeling approach taken in this paper to be robust and better than other plausible specification or prior choices. The Appendix also presents evidence against time variation in the VAR coefficients. In addition, results from the Giacomini and Rossi (2010) fluctuations test are provided and several additional analyses and details with respect to portfolio performance are provided.

5.3 | Market timing in high-volatility periods

In this subsection we present additional empirical evidence to shed more light on when our DML methods are performing well and provide some context to the existing theories of exchange rate behavior. All results are for DML with UIP which is (i) the most natural regressor choice in the context of an economic evaluation and (ii) found to perform best in the preceding subsection. We also note that results using DML with ALL REGRESSORS are very similar.

Brunnermeier, Nagel, and Pedersen (2008) document characteristic features of strategies that consider investing in high-interest-rate currencies while borrowing in low-interest-rate currencies. This delivers negatively skewed returns since the high-interest rate currencies are prone to crash risk. Similarly, Menkhoff, Sarno, Schmeling, and Schrimpf (2012) find that high-interest-rate currencies are negatively related to innovations in global FX volatility, and thus deliver low returns in times of unexpected high volatility. Inspired by papers such as these, we divide our currencies into those from countries with (on average) high interest rates (AUD, NZD, NOK, GBP) and those from countries with (on average) low interest rates (JPY, EUR, CHF, USD). As shown in the Supporting Information Appendix, the portfolio weights for these groups of currencies vary substantially over time. To investigate patterns in this variation, we regress the sum of the

⁸Della Corte et al. (2008) do not report measures of point forecasting accuracy.

portfolio weights in the high-interest-rate currencies (PW) on $FXVOL$, which is the JP Morgan index of currency volatility for the G10 countries, and find the following fitted regression line (t -statistics are in parentheses):

$$\widehat{PW} = 2.10_{(7.93)} - 10.87_{(-5.15)} FXVOL. \quad (4)$$

The message from this regression is clear cut: Our DML with UIP strategy leads to portfolios that include fewer of the high-interest-rate currencies in periods of high FX volatility, thus avoiding the crash risk associated with the carry trade strategies discussed in Brunnermeier et al. (2008) and Menkhoff et al. (2012).

There is also time variation in the economic utility produced by our DML with UIP approach relative to a random walk. If we regress the utility differences (ΔU) on $FXVOL$ (which relates specifically to currency markets), the VIX (which relates to stock markets) and $FXDIS$ (which is a measure of disagreement among professional forecasters)⁹ we find

$$\widehat{\Delta U} = 0.0254_{(-1.40)} + 0.0033_{(2.49)} FXVOL + 0.00014_{(1.21)} VIX - 0.0039_{(-1.00)} FXDIS. \quad (5)$$

This reinforces the story that DML with UIP is producing gains in utility particularly in times of high volatility in currency markets, rather than financial markets as a whole or in times of uncertainty for the professional forecasters.¹⁰

It is interesting how our findings relate to the scapegoat theory for which studies such as Fratzscher et al. (2015) and Pozzi and Sadaba (2018) find empirical support. Given our focus on developing and applying a method for out-of-sample forecasting, the suggested approach cannot be used as a direct test of the scapegoat theory. As Fratzscher et al., 2015 put it, “the theory is silent on the role of scapegoats for forecasting.” However, in the general spirit of this theory and related work on the instability of the relationship between exchange rates and fundamentals such as Bacchetta and van Wincoop, 2006 (2006, 2013), or Markiewicz (2012), we note that in times of high volatility in currency markets our DML approaches tend to include more regressors and VAR lags, which potentially reflects an intensified search for scapegoats.¹¹ And, as noted previously, we are finding a high degree of model switching and sparsity. These findings are at least consistent with this literature if not a formal test of the theoretical models advanced therein.¹²

6 | CONCLUDING REMARKS

We propose a multivariate forecasting approach for exchange rate returns that accommodates flexible dynamic model change. Our dynamic Bayesian learning approach enables us to quickly detect model changes and achieves computational feasibility by using decay factors. A major conceptual advantage of our approach over univariate models is that, by using a VAR-based approach, we obtain the input for the inherently multivariate portfolio optimization problem in a natural manner without having to rely on additional assumptions or ad hoc procedures for mapping the forecast output into portfolio weights.

We evaluate the economic value of our exchange rate forecasts in a dynamic asset allocation framework. Relying on our forecasting method, an investor achieves sizable utility gains by exploiting time-varying predictability. We establish sparsity and fast model switching as key features. Both align with the implications of the theoretical and empirical exchange rate literature.

⁹Exact definitions and data sources are given in the Supporting Information Appendix.

¹⁰The pronounced out-performance of DML strategies against the random walk (as a proxy of carry trade strategies) at the time of the subprime crisis aligns with Fratzscher (2009), who finds that currencies in which US investors held relatively large portfolio investments experienced substantially larger depreciations against the US dollar around 2008.

¹¹The Spearman correlation between the number of included regressors/VAR lags and $FXVOL$ is substantially positive, namely 0.41, for DML with UIP, and similar for other models.

¹²In line with our findings, Fratzscher et al. (2015) report out-performance of their scapegoat model against the random walk in terms of economic measures in an out-of-sample forecasting exercise, but not in terms of point forecasting accuracy.

ACKNOWLEDGMENT

We thank conference and seminar participants for comments and feedback at the 5th Conference of the International Association of Applied Econometrics (IAAE) in Montréal, the INQUIRE EUROPE Joint Spring Seminar in Windsor 2019, the Narodowy Bank Polski (NBP) Workshop on Forecasting in Warsaw 2018, the 71st Summer Meeting of the European Econometric Society (ESEM) in Cologne, the 2nd Annual Workshop on Financial Econometrics in Örebro 2018, the 1st Vienna Workshop on Economic Forecasting 2018, the Annual Conference of the German Economic Association 2018 in Freiburg, the International Conference on Macroeconomic Analysis and International Finance 2018, and the Research Seminar at the University of Innsbruck. We also appreciate helpful comments from Lucio Sarno.

OPEN RESEARCH BADGES



This article has earned an Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at [<http://qed.econ.queensu.ca/jae/datasets/beckmann001/>].

REFERENCES

- Abbate, A., & Marcellino, M. (2018). Point, interval and density forecasts of exchange rates with time varying parameter models. *Journal of the Royal Statistical Society, Series A*, 181(1), 155–179. <https://doi.org/10.1111/rssa.12273>
- Abhyankar, A., Sarno, L., & Valente, G. (2005). Exchange rates and fundamentals: Evidence on the economic value of predictability. *Journal of International Economics*, 66(2), 325–348. <https://doi.org/10.1016/j.jinteco.2004.09.003>
- Ackermann, F., Pohl, W., & Schmedders, K. (2016). Optimal and naive diversification in currency markets. *Management Science*, 63(10), 3347–3360. <https://doi.org/10.1287/mnsc.2016.2497>
- Andrews, D. W., & Monahan, J. C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, 60(4), 953–966.
- Bacchetta, P., & van Wincoop, E. (2004). A scapegoat model of exchange-rate fluctuations. *American Economic Review*, 94(2), 114–118. <https://doi.org/10.1257/0002828041301849>
- Bacchetta, P., & van Wincoop, E. (2006). Can information heterogeneity explain the exchange rate determination puzzle? *American Economic Review*, 96(3), 552–576. <https://doi.org/10.1257/aer.96.3.552>
- Bacchetta, P., & van Wincoop, E. (2013). On the unstable relationship between exchange rates and macroeconomic fundamentals. *Journal of International Economics*, 91(1), 18–26. <https://doi.org/10.1016/j.jinteco.2013.06.001>
- Bañbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian vector autoregressions. *Journal of Applied Econometrics*, 25(1), 71–92. <https://doi.org/10.1002/jae.2733>
- Berge, T. J. (2014). Forecasting disconnected exchange rates. *Journal of Applied Econometrics*, 29(5), 713–735. <https://doi.org/10.1002/jae.2350>
- Billio, M., Casarin, R., Ravazzolo, F., & van Dijk, H. (2013). Time-varying combinations of predictive densities using nonlinear filtering. *Journal of Econometrics*, 177(2), 213–232. <https://doi.org/10.1016/j.jeconom.2013.04.009>
- Brunnermeier, M. K., Nagel, S., & Pedersen, L. H. (2008). Carry trades and currency crashes. *NBER Macroeconomics Annual*, 23(1), 313–348.
- Byrne, J. P., Korobilis, D., & Ribeiro, P. J. (2016). Exchange rate predictability in a changing world. *Journal of International Money and Finance*, 62, 1–24. <https://doi.org/10.1016/j.jimonfin.2015.12.001>
- Carriero, A., Kapetanios, G., & Marcellino, M. (2009). Forecasting exchange rates with a large Bayesian VAR. *International Journal of Forecasting*, 25(2), 400–417. <https://doi.org/10.1016/j.ijforecast.2009.01.007>
- Cenesizoglu, T., & Timmermann, A. (2012). Do return prediction models add economic value? *Journal of Banking & Finance*, 36(11), 2974–2987. <https://doi.org/10.1016/j.jbankfin.2012.06.008>
- Chan, J. C., & Eisenstat, E. (2018). Bayesian model comparison for time-varying parameter vars with stochastic volatility. *Journal of Applied Econometrics*, 33(4), 509–532. <https://doi.org/10.1002/jae.2617>
- Chib, S., Nardari, F., & Shephard, N. (2006). Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics*, 134(2), 341–371. <https://doi.org/10.1016/j.jeconom.2005.06.026>
- Dangl, T., & Halling, M. (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics*, 106(1), 157–181. <https://doi.org/10.1016/j.jfineco.2012.04.003>
- Della Corte, P., Sarno, L., & Tsiakas, I. (2008). An economic evaluation of empirical exchange rate models. *Review of Financial Studies*, 22(9), 3491–3530. <https://doi.org/10.1093/rfs/hhn058>
- Della Corte, P., & Tsiakas, I. (2012). Statistical and economic methods for evaluating exchange rate predictability, *Handbook of exchange rates* (pp. 221–263). Hoboken, NJ: Wiley.
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 20(1), 253–263.
- Fratzscher, M. (2009). What explains global exchange rate movements during the financial crisis? *Journal of International Money and Finance*, 28(8), 1390–1407. <https://doi.org/10.1016/j.jimonfin.2009.08.008>
- Fratzscher, M., Rime, D., Sarno, L., & Zinna, G. (2015). The scapegoat theory of exchange rates: The first tests. *Journal of Monetary Economics*, 70(C), 1–21. <https://doi.org/10.1016/j.jmoneco.2014.09.001>

- Geweke, J., & Amisano, G. (2012). Prediction with misspecified models. *American Economic Review*, 102(3), 482–486. <https://doi.org/10.1257/aer.102.3.482>
- Giacomini, R., & Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, 25(4), 595–620. <https://doi.org/10.1002/jae.1177>
- Giannone, D., Lenza, M., & Primiceri, G. E. (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics*, 97(2), 436–451. https://doi.org/10.1162/REST_a_00483
- JP Morgan/Reuters (1996). Riskmetrics™: Technical document (4th ed.). New York, NY: JP Morgan/Reuters <https://www.msci.com/documents/10199/5915b101-4206-4ba0-ae2-3449d5c7e95a>
- Koop, G., & Korobilis, D. (2013). Large time-varying parameter VARs. *Journal of Econometrics*, 177(2), 185–198. <https://doi.org/10.1016/j.jeconom.2013.04.007>
- Kouwenberg, R., Markiewicz, A., Verhoeks, R., & Zwinkels, R. C. J. (2017). Model uncertainty and exchange rate forecasting. *Journal of Financial and Quantitative Analysis*, 52(01), 341–363. <https://doi.org/10.1017/S0022109017000011>
- Ledoit, O., & Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, 15(5), 850–859. <https://doi.org/10.1016/j.jempfin.2008.03.002>
- Li, J., Tsiakas, I., & Wang, W. (2015). Predicting exchange rates out of sample: Can economic fundamentals beat the random walk? *Journal of Financial Econometrics*, 13(2), 293–341. <https://doi.org/10.1093/jjfinec/nbu007>
- Lizardo, R. A., & Mollick, A. V. (2010). Oil price fluctuations and us dollar exchange rates. *Energy Economics*, 32(2), 399–408. <https://doi.org/10.1016/j.eneco.2009.10.005>
- Markiewicz, A. (2012). Model uncertainty and exchange rate volatility. *International Economic Review*, 53(3), 815–844. <https://doi.org/10.1111/j.1468-2354.2012.00702.x>
- Meese, R. A., & Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, 14(1), 3–24. [https://doi.org/10.1016/0022-1996\(83\)90017-X](https://doi.org/10.1016/0022-1996(83)90017-X)
- Menkhoff, L., Sarno, L., Schmeling, M., & Schrimpf, A. (2012). Carry trades and global foreign exchange volatility. *Journal of Finance*, 67(2), 681–718. <https://doi.org/10.1111/j.1468-2354.2012.00702.x>
- Pozzi, L., & Sadaba, B. (2018). Detecting scapegoat effects in the relationship between exchange rates and macroeconomic fundamentals: A new approach. *Macroeconomic Dynamics*, 2018, 1–44. <https://doi.org/10.1017/S1365100518000585>
- Rossi, B. (2006). Are exchange rates really random walks? Some evidence robust to parameter instability. *Macroeconomic Dynamics*, 10(1), 20–38. <https://doi.org/10.1017/S1365100506050085>
- Rossi, B. (2013). Exchange rate predictability. *Journal of Economic Literature*, 51(4), 1063–1119. <https://doi.org/10.1257/jel.51.4.1063>
- Sarno, L., & Valente, G. (2009). Exchange rates and fundamentals: Footloose or evolving relationship? *Journal of the European Economic Association*, 7(4), 786–830. <https://doi.org/10.1162/JEEA.2009.7.4.786>
- West, M., & Harrison, J. (1997). *Bayesian forecasting and dynamic models* (2nd ed.). Berlin, Germany: Springer. <http://www.worldcat.org/isbn/0387947256>
- Wright, J. H. (2008). Bayesian model averaging and exchange rate forecasts. *Journal of Econometrics*, 146(2), 329–341. <https://doi.org/10.1016/j.jeconom.2008.08.012>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Beckmann J, Koop G, Korobilis D, Schussler RA. Exchange rate predictability and dynamic Bayesian learning. *J Appl Econ*. 2020;35:410–421. <https://doi.org/10.1002/jae.2761>